

# Technical Report: **The Last Word** (fall 2019)

**Team:** Alvin Lo, Anu Kriti Wadhwa, Avinash Damania, Sonali Kondapalli, Sriram Hariharan

## Introduction

Endangered languages are languages that are at risk of dying out and becoming extinct. A language is classified as endangered when there are very few native speakers which have a decreasing trend over time.

Causes that may contribute to the endangerment of a language can be categorized into three main groups - physical, socio-economic and political. Physical factors such as natural disasters, violent wars, and genocide which are powerful or prevalent enough to wipe out populations of native language speakers.

Socio-economic reasons include urbanization, where youth are moving from native areas to cities and bringing up their families there instead. Resultantly, the following generation would have low incentives to learn their native languages which are not used in the urban areas in which they reside. Language endangerment is also caused by cultural and economic marginalization, as youth may once again favor learning more popular and arguably, more functional languages for purposes such as career advancement. Inter-marriage between people from different native group languages could contribute to language endangerment. This is because, for convenience in communication among the family, any children they may have may only learn one of the two languages their parents speak.

An example of political marginalization could be when a government of a large nation adopts a singular official language. This causes smaller areas to have a language shift towards the official language for functional purposes, endangering their native languages. Political repression happens when leadership within nations promotes having a singular national culture that is focussed around one language. This immediately limits opportunities for people in other language groups, forcing them to shift from their native language to the promoted one. These groups tend to be minorities in the nations and may be forced to resettle, or even break away from their families. Over time, this results in potential language endangerment as well.

## Motivation

Similar to endangered species, there are arguments which claim that the extinction of languages is due to evolution, and the idea of 'survival of the fittest'. Languages that prevail today are the most utilized as they are functional in communication among large groups of people, increasing productivity, efficiency, and convenience. However, with the loss of some languages, we would be losing not only their semantics, such as phrases, expressions or grammatical rules but also the wealth of meaning they hold in understanding the culture in

which they are used. These languages provide cultural identity and communal belonging. Without them, there is loss of information, traditions, songs, anecdotes, which may be historical, of medicinal value, or anthropological significance.

There are institutions and nonprofit organizations around the world that are dedicated to saving these languages. There are also multiple efforts in documenting and recording as much as possible to ensure that language-specific records remain and can be understood in the unfortunate scenario that the language becomes extinct.

There are also efforts being made in language revitalization, where endangered languages are being preserved by being taught to children, encouraging them to become fluent in the tongue.

Lastly, technology contributes to these efforts as well. Media such as digital classrooms, podcasts, audio recordings, phone applications and computer programs made in these endangered languages promote their use.

There is little awareness of these dying languages outside of the relevant communities, such as researchers in the field, charities for the cause, or native speakers. The resources we found on the topic tend to be either overwhelming data sets which are inaccessible to most, or dense publications which are difficult to interpret for most unless they are familiar with the terminologies. With our website, we hope to create a resource to raise awareness on the issue and provide relevant information to users in a simple, easy-to-comprehend fashion. In addition to parsing the data sets and presenting them in a more concise manner, we also hope to provide additional resources for those who would like to learn more. These resources could be general information on the issue of endangered languages or language-specific.

*Pariona, Amber. "Why Do We Need To Save Dying Languages?" WorldAtlas, 8 Aug. 2017, <https://www.worldatlas.com/articles/why-do-we-need-to-save-dying-languages.html>.*

## User stories

1. Add a non-default favicon tab, so it is easier to navigate to the page when multiple tabs are open.

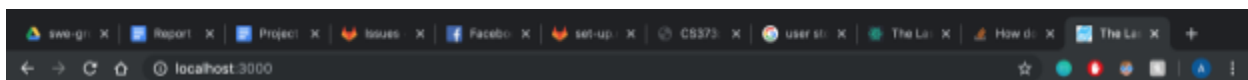




Figure 1: Favicon when multiple tabs are open

As Figure 1 shows above, we changed our favicon tab from the generic  to , to make it more easily identifiable by users who may have multiple tabs open.

2. Have a more attractive splash page

“To speak a language is to take on a  
world, a culture.”

- Frantz Fanon

There are approximately **7000** languages in the world. Almost **40%** of them are classified as Threatened.

Explore our site to learn more about these languages and what you can do to help save them!

Figure 2a: Previous Splash Page



“To speak a **language**  
is to  
take on a **world**, a  
**culture.**”

- Frantz Fanon

Almost 40% of the languages in the world are  
classified as “threatened”.

If these languages were to become extinct,

© Copyright 2019 SWE Group 18

Figure 2b: Updated Splash Page

We updated the splash page to make it more attractive, by adding more color, images, and emphasis on different portions. Figure 2a is what we had previously, while Figure 2b reflects these changes.

3. We added information for the main spoken languages in each country, as in Figure 3.

# Canada

**Main Languages:** English, French

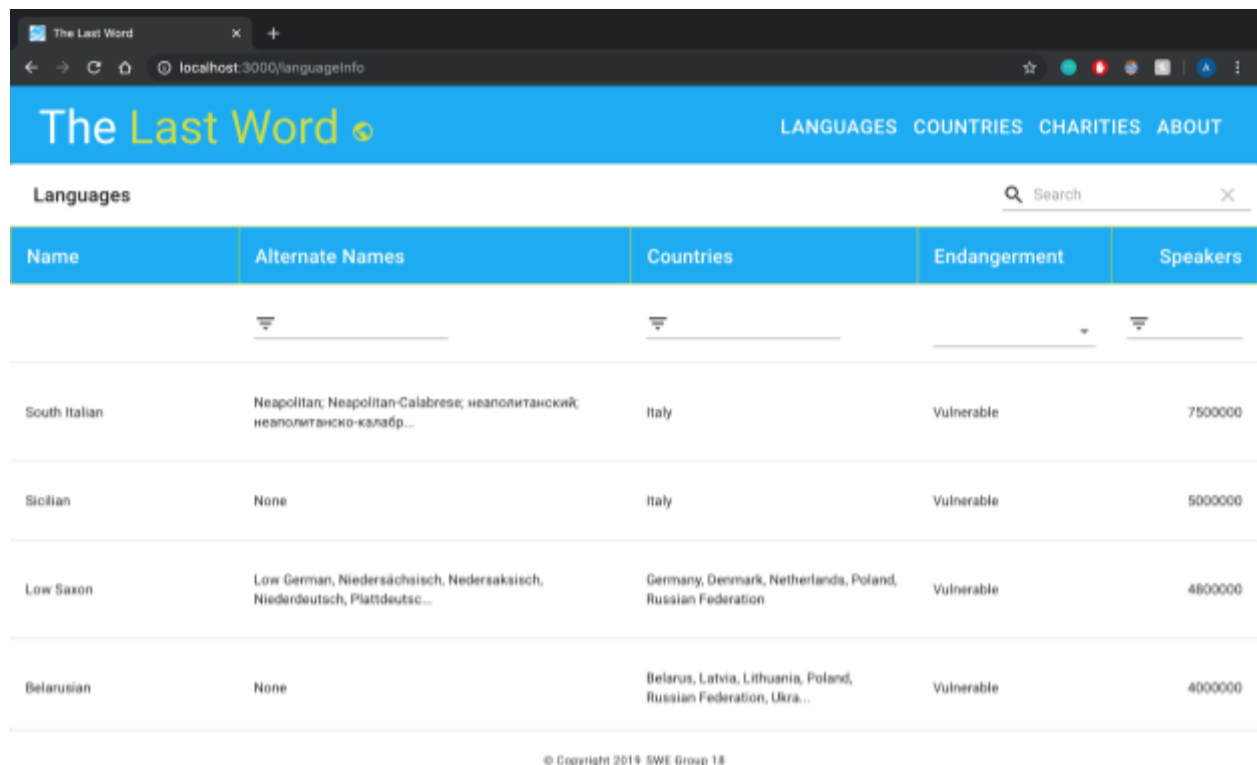
Figure 3: Main Languages of Country

This change can be seen on the 'countries' page.

4. Cards on about page not clickable

We did make them not clickable when we were still using media cards. However, we updated the page to have a tabular format, and this user story no longer applied to our website layout.

This is seen in Figure 4 below:



Name	Alternate Names	Countries	Endangerment	Speakers
South Italian	Neapolitan; Neapolitan-Calabrese; неаполитанский; неаполитанско-калабр...	Italy	Vulnerable	7500000
Sicilian	None	Italy	Vulnerable	5000000
Low Saxon	Low German, Niedersächsisch, Nedersaksisch, Niederdeutsch, Plattdeutsc...	Germany, Denmark, Netherlands, Poland, Russian Federation	Vulnerable	4800000
Belarusian	None	Belarus, Latvia, Lithuania, Poland, Russian Federation, Ukra...	Vulnerable	4000000

© Copyright 2019 SWE Group 18

Figure 4: Tabular format

## Tools

### Material table

We used a react library, called material-table, to display the sets of data for our models. We chose this library primarily because it worked with our existing react code. It is fully customizable, which made it convenient to reformat to fit the data sets we needed. Additionally, it allows for easy navigation across each table, provides tools to manipulate and query the data, and lastly would allow for a corresponding visualization for future parts of the project.

We implemented **search** using properties of the material table library as well. We chose to do this as it was the most convenient method to implement search with the existing code we had, in a manner that allowed it to be customizable should we change the representation of our data in the future.

### Cheerio/Beautiful Soup

Initially, we used Beautiful Soup to scrape Ethnologue. However, in the past two weeks, they updated their paywall and we were no longer able to scrape from the site. In light of this, we looked for new sources of data for endangered languages which are the following:

The new country API:

1. <https://datahelpdesk.worldbank.org/knowledgebase/articles/898590-country-api-queries?fbclid=IwAR2lu-121C1yJ9cCDFzslq5CZfno0XLHa-TBbFH1phfXAIJhV9HFxeiAJ3o>
2. [http://api.worldbank.org/v2/country?format=json&fbclid=IwAR2KLsh6BxpEP0vKMh8Hep7lWiQ1i1Do2FaPHOu4F\\_wY7RrU0wtbnnnt-rU](http://api.worldbank.org/v2/country?format=json&fbclid=IwAR2KLsh6BxpEP0vKMh8Hep7lWiQ1i1Do2FaPHOu4F_wY7RrU0wtbnnnt-rU)

The new language API:

1. [https://www.theguardian.com/news/datablog/2011/apr/15/language-extinct-endangered?fbclid=IwAR1MBCFP8xFgqloPbs3mmY-1fPGK2YshCcfozgDFiH\\_FuKkQ4gLYf1KDDPg](https://www.theguardian.com/news/datablog/2011/apr/15/language-extinct-endangered?fbclid=IwAR1MBCFP8xFgqloPbs3mmY-1fPGK2YshCcfozgDFiH_FuKkQ4gLYf1KDDPg)

For our Charities, we did a scrape for charities related to the keyword “language” on charity navigator and scraped all the data for

We used cheerio and beautiful soup to scrape these sites for the information we needed. They

### Amazon Relational Database Service (RDS)

We chose to use Amazon RDS as it made our backend database easy to set up, operate and scale for possible future needs, all in the cloud. We used only the free tier of the service.